Using Interpretable Machine Learning for Analyzing Acoustic Startle Response in Rodents from Acoustic Emissions Data

MIR MOHAMMAD SHAMSZADEH,
SHIVASHANKAR PERUVAZHUTHI,
LAURA AGEE, MICHAEL DREW
and SALVATORE SALAMONE

ABSTRACT

Rodent behavior studies in neuroscience and psychology investigate various sensory and motor responses. One such study examines the acoustic startle response (ASR), an unconditional reflex to an unexpected and intense acoustic stimulus characterized by a rapid contraction of facial and skeletal muscles. These contractions generate subtle mechanical forces that propagate through the surface that the animal is in contact with, resulting in acoustic emission (AE) signals. The AE signals travel through the plate as guided ultrasonic waves, carrying information about the animal's motor response to ASR and its physiology, behavior, and underlying psychological state. Various ASR experiments were conducted to assess animal's reactions to sudden auditory stimuli. The experiments involved three strains of mice (129S6/SvEv, CBA/CaJ, and C57BL/6J) in an open field equipped with AE sensors. The AE data were collected as discrete wavepackets (AE hits) using an amplitude threshold-based acquisition system, with each hit containing signal features that characterize the corresponding movement or behavior.

We propose the use of the Explainable Boosting Machine (EBM) to analyze AE data generated by the animal during movement. Our goal is to distinguish ASR-related AE hits from normal AE activity, aiming to analyze the characteristics of each hit to confirm its association with ASR events. This classification is crucial not just for the detection of ASR occurrences but for analyzing the intensity and variability of startle responses across different conditions and strains. The dataset consists of AE-derived features (e.g., amplitude and frequency) combined with biometric factors (e.g., weight, sex) and environmental variables (e.g., lighting conditions, stress exposure). By leveraging EBM's interpretability, this study identifies key factors influencing ASR and provides a transparent, data-driven approach for analyzing ASR in rodents. EBM provides an effective means to understand complex animal-structure interaction and to analyze the dynamic interplay between body movements and surrounding structures, revealing their mutual influences and shedding light on intricate biomechanical patterns.

Mir Mohammad Shamszadeh[1], Shivashankar Peruvazhuthi[1], Laura Agee[2], Michael Drew[2], Salvatore Salamone[1]

[1]Smart Structures Research Laboratory, Department of Civil, Architectural, and Environmental Engineering, University of Texas at Austin, Austin, Texas, U.S.A.

[2]Department of Neuroscience, University of Texas at Austin, Austin, Texas, U.S.A.

INTRODUCTION

This study focuses on analyzing acoustic emission (AE) signals transmitted through the surface beneath rodents in response to auditory stimuli, aiming to identify patterns associated with the acoustic startle response (ASR). ASR is defined as a simple response characterized by a rapid contraction of facial and skeletal muscles following an unexpected and intense acoustic stimulus [1]. It is used to investigate behavioral and neurological responses to sensory stimuli under controlled experimental conditions. In a typical ASR experiment, a sudden loud acoustic stimulus is presented to a rodent, and the resulting rapid, involuntary response is measured to assess sensory-motor reactivity and neurological function. The restrictive testing environments using commercial startle chambers and stabilimeters are intentionally designed to minimize voluntary movements, because the sensors they rely on—such as load cells, accelerometers, or force transducers—cannot reliably separate voluntary motion from true startle responses. By constraining the animal's movement, these setups aim to ensure that sensor outputs reflect only the ASR. However, this restriction can interfere with the animal's natural behavior and elevate stress or anxiety levels, ultimately affecting the validity of the measured response.

Among laboratory setups, the open field environment allows rodents to behave more naturally, closely resembling conditions in the wild. Measuring the ASR in this setting yields results that are more representative of real-world behavior and therefore more applicable to animals in natural environments. However, using traditional sensors or the piezoelectric plate may not yield distinguishable ASR output from the animal's natural movement and behavior. To address this, ASR is measured by recording AE signals generated by the animal's movements in a standard-dimension open field used [2,3]. Muscle contractions exert subtle forces on the surface beneath the animal—such as a metallic plate—producing AEs that propagate through the underlying plate. In plate-like structures with parallel free boundaries, AEs generate Lamb waves and shear horizontal (SH) waves at ultrasonic frequencies. Both Lamb and SH waves exhibit multiple modes due to the boundary conditions and thickness of the plate. Lamb modes arise from the superposition of multiple boundary reflections of bulk longitudinal and shear vertical (SV) waves [4], whereas SH modes form from the confinement and reflection of in-plane shear waves (bulk SH waves) between the plate's free surfaces. The waves can then be detected with highly sensitive AE sensors and recorded with a high time- and voltage-resolution data acquisition system.

The acquired AE signals contain information about the movement or behavior of the rodent that caused them. Even without an auditory stimulus, the animal's movements can still generate AEs through subtle mechanical interactions with the surface. By analyzing the features of the AE signals throughout the test, we can determine which characteristics effectively distinguish normal activities (i.e., non-ASRs) from ASRs. Identifying such features enables the development of automated methods for detecting ASRs. Moreover, knowing which AE features influence the classification, and how they do so, is important for uncovering the physiological or behavioral signatures of the underlying responses, guiding both interpretation and future experimental design. This understanding also enables the study of how environmental conditions—such as lighting, stress exposure, or other experimental manipulations—modulate the acoustic signatures associated with different behaviors.

EXPERIMENTAL SETUP FOR ACOUSTIC EMISSIONS ACQUISITION

The experimental setup is displayed in Figure 1. On the right of the figure, a metallic floor that enables guided wave propagation is shown. The edges were covered with damping tape to eliminate edge wave reflections. The enclosure was suspended using four lab stands to prevent any contact with the plate. Five R15i sensors (Physical Acoustics Corporation), operating in the 50-400 kHz range and equipped with integrated preamplifiers providing 40 dB gain, were attached to the bottom surface to detect AE hits. The Physical Acoustics Corporation's Express-8 AE board, housed in the Micro-II Express chassis was used as the data acquisition system. The signals are acquired and stored by the system through the amplitude threshold-triggered acquisition method. The system initiates acquisition only when the sensor signal exceeds a fixed threshold of 35 dB. Signals are sampled at 5 MHz and bandpass filtered between 20 kHz and 400 kHz prior to storage and further processing.

Ten-millisecond single-frequency tone bursts, used as startle stimuli, were played from a MacBook Air, amplified using an Aiyima A07 power amplifier, and delivered through a Timpano TPT-ST25 tweeter positioned above the open field. The tweeter, placed over the open field, is rated for frequencies up to 20 kHz. This audio setup can deliver startle tones at levels up to 110 dB for 20 kHz and 130 dB for other lower frequencies. The desired sound levels were obtained by choosing an optimum amplifier gain and then adjusting the amplitude of the audio files. Tone bursts were presented multiple times during the tests. The audio tone used as the startle stimulus was also acquired as an AE signal. The laptop's audio output was split into two paths: one was amplified using a 2/4/6 preamplifier and recorded as an AE, while the other was amplified and sent to the tweeter for sound delivery. Capturing the audio tone as an AE enables precise determination of the exact moment the stimulus is delivered.

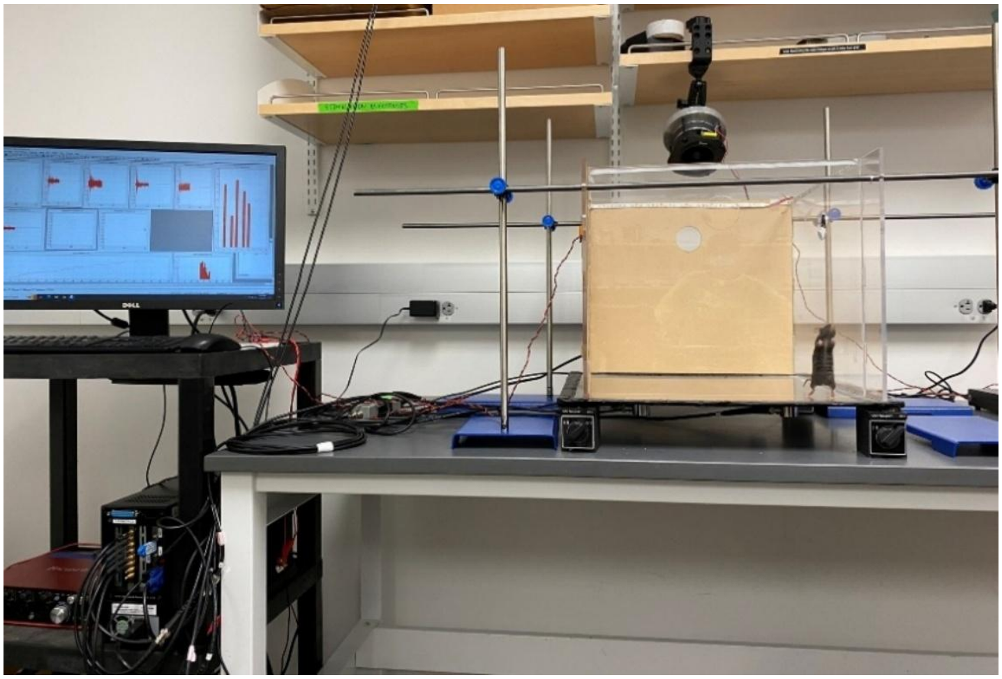

Figure 1. The experimental setup

DATASET ACQUISITION AND CHARACTERIZATION

In this section, we describe the dataset obtained through our experimental campaign. Multiple experiments were conducted on rodents to investigate the effects of various factors on the ASR AE hits. Mice from three strains—129S6/EvSv, CBA/CaJ, and C57BL/6J—including both males and females, were tested across different tone frequencies under varying conditions of lighting and stress exposure, with animal's weight recorded for each subject. Lighting conditions included light and dark environments to examine the effect of light v. dark condition on the startle responses. Stress exposure was defined by whether the mice underwent fear conditioning with ten 1 mA shocks over 11 minutes, 60 minutes prior to the test. The unstressed group was placed in the same environment without shocks.

Figure 2 shows the absolute energy of each AE hit against the corresponding time-of-hit over the 10-minute test duration of a male CBA mouse. Non-ASR AE hits are denoted by white colors. All AE hits occurring within a 150 ms time span from the onset of a startle tone are classified as the ASR AE hits for that specific tone. These ASR AE hits are highlighted in various non-white colors, with each color representing a specific sensor. In many cases, the waves generated by the animal are detected by more than one sensor, even when the animal is not near those sensors. This results in AE hits from multiple sensors being recorded for a single ASR event.

After conducting all tests across mouse strains, stress exposures, and lighting conditions, the combined dataset comprises 325,695 AE hits. Each AE hit represents an acoustic event associated with the physical movements/reflexive responses of a rodent, captured by the acoustic sensors. Each AE hit is represented by a rich set of features, including signal characteristics (such as root mean square (RMS), duration, amplitude, peak frequency, and average frequency), and contextual metadata (including mouse strain, sex, weight, lighting condition, and stress exposure).

Figure 3 shows the correlation matrix for the signal features. Additionally, each hit is labeled with an ASR label indicating whether the event corresponds to an ASR or not, according to the procedure defined above, for classification purposes. This structured dataset serves as the foundation for training machine learning models aimed at classifying each AE hit as either an ASR-related or non-ASR event, based on signal and contextual features. By employing an interpretable machine learning approach, we can identify which features most strongly influence the classification and in what manner. This, in turn, enables us to associate specific AE signal characteristics with distinct behavioral responses in mice.

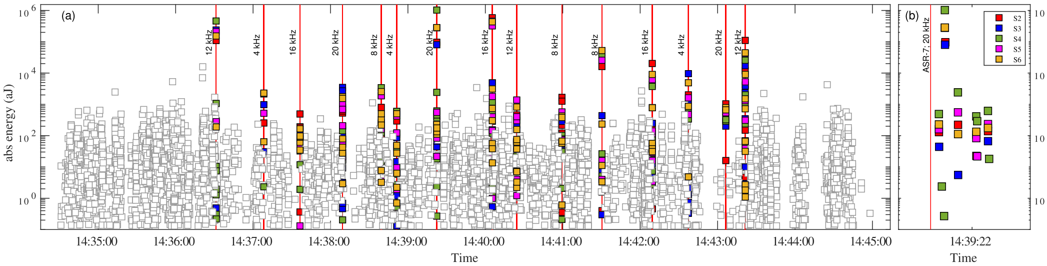

Figure 2. (a) Absolute energy of AE hits from all five R15i sensors for a male CBA mouse, and (b) close-up view of the seventh ASR. Red vertical lines mark when startle tones occurred. White boxes show non-ASR hits; colored boxes show ASR hits, with color showing which sensor detected the hit.

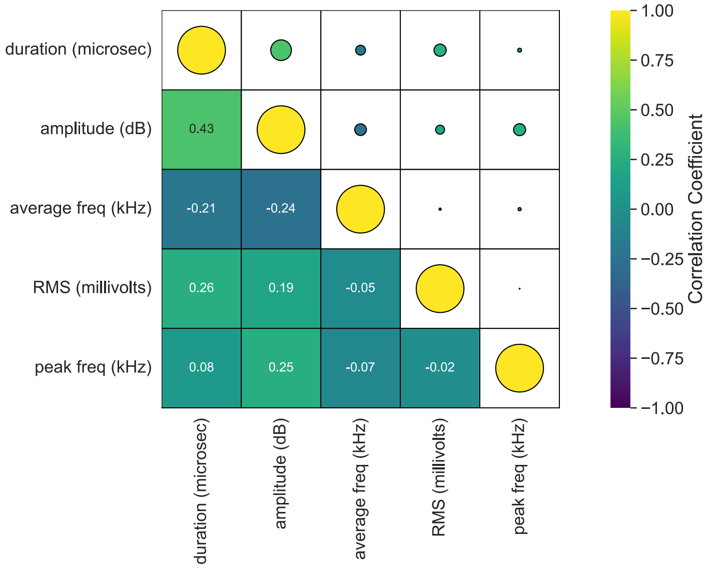

Figure 3. Correlation matrix for the signal features

EXPLAINABLE BOOSTING MACHINE

Explainable Boosting Machine (EBM) [5,6] was chosen for analyzing the dataset due to its inherent interpretability. It is a glass-box model that predicts the response variable as an additive combination of nonlinear functions applied to individual features, along with a limited number of nonlinear pairwise interaction terms. It achieves high predictive accuracy while preserving model transparency and interpretability.

For a binary classification problem, EBM models the predicted response, $\hat{y}$, in an additive form by expressing the outcome as a sum of univariate and pairwise interaction nonlinear functions:

$$\hat{y} = g(\mathbb{E}[y|\boldsymbol{x}]) = f_0 + \sum_{j=1}^{d} f_j(x_j) + \sum_{\substack{1\le j<k\le d \\ (j,k)\in\mathfrak{T}}} f_{jk}(x_j, x_k) \tag{1}$$

where f_0 is the intercept term, $f_j(x_j)$ is a univariate shape function which represents the main effect of the feature x_j on the response variable, $\mathfrak{T}$ represents the set of feature pairs, $f_{jk}(x_j, x_k)$ is a bivariate shape function which represents the pairwise interaction between features x_j and x_k on the response variable, and $g(.)$ is the logit link function in a classification problem. In the EBM training process, the model incrementally learns and refines its predictions through an iterative boosting framework.

EBM is interpretable because each univariate shape function's relationship with the response variable can be visualized through a plot of $f_j(x_j)$ versus x_j. Moreover, a pairwise interaction can be rendered as a heatmap of $f_{jk}(x_j, x_k)$ on the x_j-x_k plane.

Environmental and biological features such as weight, sex, lighting condition, and stress exposure are treated as contextual metadata—system parameters that moderate the relationship between AE signal features and the classification outcome. These contextual features are not assigned univariate shape functions but appear only through interaction terms with AE signal features. In contrast, AE signal characteristics (e.g., amplitude, energy, frequency components), their univariate effects, and their interactions—both among themselves and with contextual features—form the primary features and directly contribute to the classification between an ASR AE hit and non-ASR AE hit. This structure reflects the assumption that while AE signals capture immediate behavioral responses, the contextual features shape how that response presents under different experimental conditions.

RESULTS AND DISCUSSION

After training the EBM model for the classification task, the influence of each feature can be quantified using its mean absolute log-odds score, which reflects its relative importance compared to other features. In addition, the model provides interpretable visualizations of the learned univariate and bivariate shape functions, showing how each univariate or bivariate feature affects the prediction. It should be noted that the purpose of classification in this study is to analyze patterns within the data, rather than to develop a predictive model for deployment.

Feature Importance

The feature importance plot (Figure 4) ranks the overall contribution of each feature to the model’s prediction using the mean absolute contribution of each term. Among the features included, duration and RMS emerged as the most influential in distinguishing between ASR and non-ASR AE hits. Following these, three pairwise interactions—duration and amplitude, amplitude and average frequency, and duration and average frequency—had the highest contributions. Duration of the AE hit corresponds to how long the physical response lasted or how long the force was applied, and RMS represents the average intensity of the AE signal, providing a measure of the sustained intensity of the movement. Amplitude reflects the maximum AE signal excursion during an event, indicating a stronger force exerted on the surface. Frequency is associated with the speed of the movement, with higher frequencies indicating more rapid physical responses.

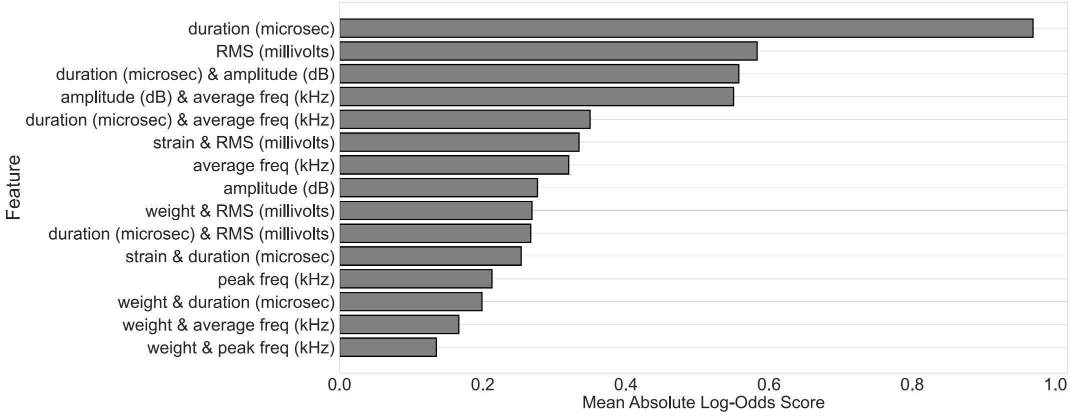

Figure 4. Global feature importance.

Among the contextual features, strain of the mice was the most significant factor influencing classification, followed by weight. In contrast, lighting condition and stress exposure did not appear among the top 15 most important features. These findings suggest that the strain of mice has a clear influence on ASR behavior, indicating that different strains respond differently to the same auditory stimulus. Body weight also plays a role, likely because it affects the characteristics of the acoustic emission signal generated during the response. In contrast, lighting condition (light vs. dark) and stress exposure show little impact on ASR occurrence, suggesting that the ASR in rodents is largely unaffected by these environmental conditions under the tested setup.

Model-Derived Effects of AE Signal and Contextual Features

The functional form of each shape function in the EBM model can be visualized separately. Figure 5 shows the univariate shape function for duration, where the y-axis represents the log-odds score, indicating the effect of duration on the model's prediction, with higher scores corresponding to a greater probability of classifying an event as ASR. Overall, longer durations are associated with an increased likelihood of non-ASR classification, suggesting that ASR events typically involve shorter-duration AE hits—i.e., brief force exertion by the rodent. However, extremely short durations, as observed on the far left of the plot, are also linked to non-ASR events, possibly reflecting background noise or unrelated minor movements. The red bands represent uncertainty intervals, with wider intervals indicating lower confidence in the learned pattern in those regions. Similar plots can be generated for other main effects. For instance,

Figure 6 presents the univariate shape function for RMS, showing that higher RMS values correspond to a greater likelihood of ASR classification, which is consistent with the expectation that ASR events involve higher-intensity AE signals.

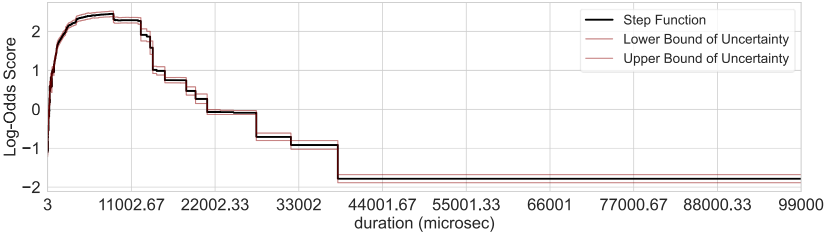

Figure 5. Univariate shape function plot for duration.

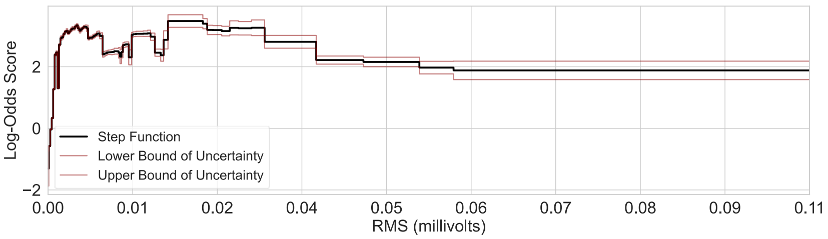

Figure 6. Univariate shape function plot for RMS.

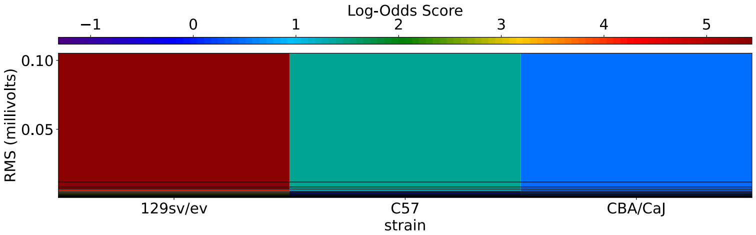

Figure 7. Interaction heatmap of RMS and mouse strain for ASR classification.

Rendering $f_{jk}(x_j, x_k)$ against the (x_j, x_k) feature pairs—where one of the features is a contextual variable (e.g., strain)—using interaction heatmaps provides insight into how different contextual features influence the effect of an AE signal feature. Figure 7 presents the pairwise interaction between mouse strain and the RMS of the AE signal. As shown, given the same RMS value, the model assigns a higher probability of ASR to events from the 129S6/SvEv strain. This suggests that high RMS values in this strain are more indicative of a startle response than in the CBA/CaJ and C57BL/6J strains.

CONCLUSIONS

Acoustic emission (AE) sensors were used in multiple experiments to detect and quantify acoustic startle responses (ASR) in freely moving mice. Movements and startles of the animal generate Lamb waves that propagate through the surface. These waves carry rich information about the underlying physical action. By analyzing the AE signals, one can infer and quantify the nature of the behavioral response. To interpret these AE patterns, Explainable Boosting Machine was employed, which learns additive shape functions and pairwise interactions, allowing transparent insights into how AE signal features contribute to ASR classification. Moreover, the inclusion of contextual features, such as mouse strain or lighting condition, enables the model to capture modulation of behavioral responses by biological and environmental factors.

REFERENCES

1. Koch M, Schnitzler HU. The acoustic startle response in rats - Circuits mediating evocation, inhibition and potentiation. Behavioural Brain Research 1997;89:35–49. https://doi.org/10.1016/S0166-4328(97)02296-1.
2. Shivashankar P, Nocera N, Livadiotis S, Mozaffar S, Drew MR, Salamone S. Methodology of using acoustic emissions for enhancing rodent behavioral analysis. Ultrasonics 2024;136:107170. https://doi.org/10.1016/J.ULTRAS.2023.107170.
3. Shivashankar P, Agee LA, Wilson N, Salamone S, Drew MR. High-fidelity detection and quantification of acoustic startle response in freely moving mice in an open field. J Neurosci Methods 2025;422:110497. https://doi.org/10.1016/J.JNEUMETH.2025.110497.
4. Rose JL. Ultrasonic Guided Waves in Solid Media. Ultrasonic Guided Waves in Solid Media 2014;9781107048959:1–512. https://doi.org/10.1017/CBO9781107273610.
5. Lou Y, Caruana R, Gehrke J, Hooker G. Accurate intelligible models with pairwise interactions. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2013;Part F128815:623–31. https://doi.org/10.1145/2487575.2487579.
6. Nori H, Jenkins S, Koch P, Caruana R. InterpretML: A Unified Framework for Machine Learning Interpretability 2019.